



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo

**Citation for published version:**

Cockram, CA, Filatenkova, M, Danos, V, El Karoui, M & Leach, DRF 2015, 'Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo', *Proceedings of the National Academy of Sciences (PNAS)*, vol. 112, no. 34, E4735-E4742.  
<https://doi.org/10.1073/pnas.1424269112>

**Digital Object Identifier (DOI):**

[10.1073/pnas.1424269112](https://doi.org/10.1073/pnas.1424269112)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Proceedings of the National Academy of Sciences (PNAS)

**Publisher Rights Statement:**

Freely available online through the PNAS open access option.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo

Charlotte A. Cockram<sup>a</sup>, Milana Filatenkova<sup>b,c</sup>, Vincent Danos<sup>b,c</sup>, Meriem El Karoui<sup>a,c,1</sup>, and David R. F. Leach<sup>a,1</sup>

<sup>a</sup>Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom; <sup>b</sup>Life Sciences Institute, School of Informatics, University of Edinburgh, Edinburgh EH8 9LE, United Kingdom; and <sup>c</sup>SynthSys, Centre for Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom

Edited by Gerald R. Smith, Fred Hutchinson Cancer Research Center, Seattle, WA, and accepted by the Editorial Board July 15, 2015 (received for review December 18, 2014)

**Understanding molecular mechanisms in the context of living cells requires the development of new methods of in vivo biochemical analysis to complement established in vitro biochemistry. A critically important molecular mechanism is genetic recombination, required for the beneficial reassortment of genetic information and for DNA double-strand break repair (DSBR). Central to recombination is the RecA (Rad51) protein that assembles into a spiral filament on DNA and mediates genetic exchange. Here we have developed a method that combines chromatin immunoprecipitation with next-generation sequencing (ChIP-Seq) and mathematical modeling to quantify RecA protein binding during the active repair of a single DSB in the chromosome of *Escherichia coli*. We have used quantitative genomic analysis to infer the key in vivo molecular parameters governing RecA loading by the helicase/nuclease RecBCD at recombination hot-spots, known as Chi. Our genomic analysis has also revealed that DSBR at the *lacZ* locus causes a second RecBCD-mediated DSBR event to occur in the terminus region of the chromosome, over 1 Mb away.**

homologous recombination | mechanistic modelling | DNA repair | RecA | RecBCD

**D**NA double-strand break repair (DSBR) is essential for cell survival and repair-deficient cells are highly sensitive to chromosome breakage. In *Escherichia coli*, a single unrepaired DNA DSB per replication cycle is lethal, illustrating the critical nature of the repair reaction (1). DSBR in *E. coli* is mediated by homologous recombination, which relies on the RecA protein to efficiently recognize DNA sequence identity between two molecules. RecA homologs are widely conserved from bacteriophages to mammals, where they are known as the Rad51 proteins (2). The RecA protein plays its central role by binding single-stranded DNA (ssDNA) to form a presynaptic filament that searches for a homologous double-stranded DNA (dsDNA) donor from which to repair. It then catalyzes a strand-exchange reaction to form a joint molecule (3), which is stabilized by the branch migration activities of the RecG and RuvAB proteins (4). The joint molecule is then resolved by cleavage at its four-way Holliday junction by the nuclease activity of RuvABC (5, 6).

RecA binding at the site of a DSB is dependent on the activity of the RecBCD enzyme (Fig. 1A). RecBCD is a helicase-nuclease that binds to dsDNA ends, then separates and unwinds the two DNA strands using the helicase activities of the RecB and RecD subunits (see refs. 7 and 8 for recent reviews). RecD is the faster motor of the two and this consequently results in the formation of a ssDNA loop ahead of RecB (Loop 1 in Fig. 1A) (9). As the enzyme translocates along dsDNA, the 3'-terminated strand is continually passed through the Chi-scanning site thought to be located in the RecC protein (10). When a Chi sequence (the octamer 5'-GCTGGTGG-3') enters this recognition domain, the RecD motor is disengaged and the 3' strand continues to be unwound by RecB. Under in vitro conditions, where the con-

centration of magnesium exceeds that of ATP, the 3' end (unwound by RecB) is rapidly digested before Chi recognition, whereas the 5' end (unwound by RecD) is intermittently cleaved (11, 12). After Chi recognition the 3' end is no longer cleaved but the nuclease domain of RecB continues to degrade the 5' end as it exits the enzyme (11, 12). Under in vitro conditions where the concentration of ATP exceeds that of magnesium, unwinding takes place but the only site of cleavage detected is ~5 nucleotides 3' of the Chi sequence (13, 14). Because the RecB motor continues to operate while the RecD motor is disengaged, Loop 1 is converted to a second loop located between the RecB and RecC subunits or to a tail upon release of the Chi sequence from its recognition site. We therefore describe this single-stranded region as Loop/Tail 2 in Fig. 1A. After the whole of Loop 1 is converted to Loop/Tail 2, this second single-stranded region continues to grow as long as the RecB subunit unwinds the dsDNA. The RecBCD enzyme enables RecA protein to load on to Loop/Tail 2 to generate the presynaptic filament necessary to search for homology and initiate strand-exchange (15). Finally, the RecBCD enzyme stops translocation and disassembles as it dissociates from the DNA, releasing a DNA-free RecC subunit (16).

Our understanding of the action of RecBCD and RecA has been the result of more than 40 years of genetic analysis and

## Significance

**Maintaining genomic integrity is crucial for cell survival. In *Escherichia coli*, RecA-mediated homologous recombination plays an essential role in the repair of DNA double-strand breaks (DSBs). A greater understanding of the mechanism of homologous recombination requires quantitative analysis of genomic studies in live cells. We have developed a novel method that is able to capture these interactions on a genome-wide scale by combining ChIP-seq and mathematical modeling to interpret the patterns of RecA-DNA interaction during DSB repair (DSBR). This genomic analysis has also revealed unexpected RecA binding in the terminus region of the chromosome, consistent with a second DSBR event (at a distance of 1 Mb) that is indirectly caused by the first DSBR event induced at the *lacZ*.**

Author contributions: C.A.C., M.E.K., and D.R.F.L. designed research; C.A.C. and M.F. performed research; C.A.C., M.F., V.D., M.E.K., and D.R.F.L. analyzed data; and C.A.C., V.D., M.E.K., and D.R.F.L. wrote the paper.

The authors declare no conflict of interest.

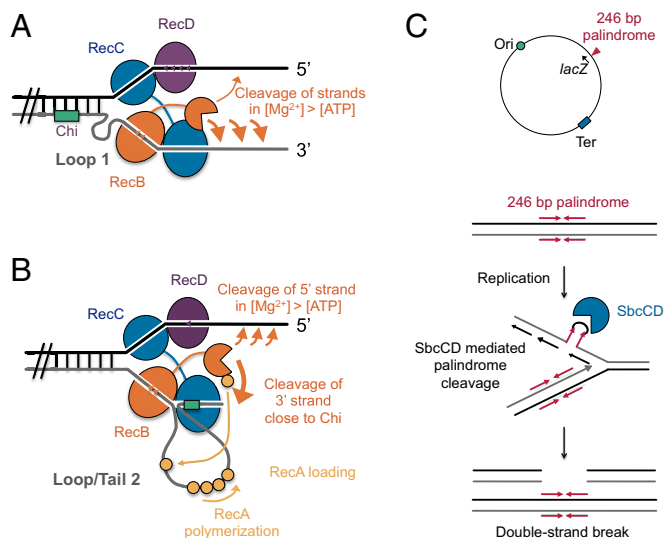
This article is a PNAS Direct Submission. G.R.S. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo) (accession no. GSE71249).

<sup>1</sup>To whom correspondence may be addressed. Email: Meriem.Elkaroui@ed.ac.uk or D.Leach@ed.ac.uk.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1424269112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1424269112/-DCSupplemental).



**Fig. 1.** DSBR in *E. coli*. (A and B) Schematic representation of DSB processing by the RecBCD complex. (A) Before Chi recognition, both the RecB and RecD motors progress along the DNA. RecD is the faster motor and as a result a loop of ssDNA (Loop 1) is formed ahead of the slower RecB motor. The 3' ssDNA strand is scanned for the Chi sequence by the RecC protein. (B) After Chi recognition, RecBCD likely undergoes a conformational change so that only the RecB motor is engaged. The RecA protein is recruited by the RecB nuclease domain and loaded onto the ssDNA loop generated by RecB unwinding to promote RecA nucleoprotein filament formation. In this schematic representation, the Chi site is shown held in its recognition site. However, the Chi site will be released either by disassembly of the RecBCD complex or at some point before this and the second single-stranded region on the 3' terminating strand will be converted from a loop to a tail. Therefore, this region is denoted Loop/Tail 2. The mathematical model described in *SI Appendix* does not depend on the ATP/magnesium dependent differential cleavage of DNA strands (7, 8), nor does it depend on the precise time that the 3' end is released from the complex following Chi recognition. (C) The hairpin endonuclease SbcCD is used to cleave a 246-bp interrupted palindrome inserted in the *lacZ* gene of the *E. coli* chromosome. Cleavage of this DNA hairpin results in the generation of a site-specific DSB on only one of a pair of replicating sister chromosomes, thus leaving an intact sister chromosome to serve as a template for repair by homologous recombination.

biochemical investigation of these purified proteins in vitro. However, relatively little is known about their activities on the genomic scale. To investigate these reactions in vivo, we have used RecA chromatin immunoprecipitation with next-generation sequencing (ChIP-Seq) in an experimental system that allows us to introduce a single and fully repairable DSB into the chromosome of *E. coli* (1). Because DSBR by homologous recombination normally involves the repair of a broken chromosome by copying the information on an unbroken sister chromosome, our laboratory has previously developed a procedure for the cleavage of only one copy of two genetically identical sister chromosomes (1). We have made use of the observation that the hairpin nuclease SbcCD specifically cleaves only one of the two sister chromosomes following DNA replication through a 246-bp interrupted palindrome to generate a two-ended DSB (1). As shown in Fig. 1B, this break is fully repairable and we have shown that recombination-proficient cells suffer very little loss of fitness in repairing such breaks (17).

Here we investigate in vivo and in a quantitative manner the first steps of DSBR: because the outcome of RecBCD action is understood to be the loading of RecA on DNA in a Chi-dependent manner, we use RecA-ChIP to reveal the consequences of RecBCD action on a genomic scale during DSBR. Analyses of most ChIP-Seq datasets focus on the identification of regions of significant enrichment of a given protein but do not take into account the

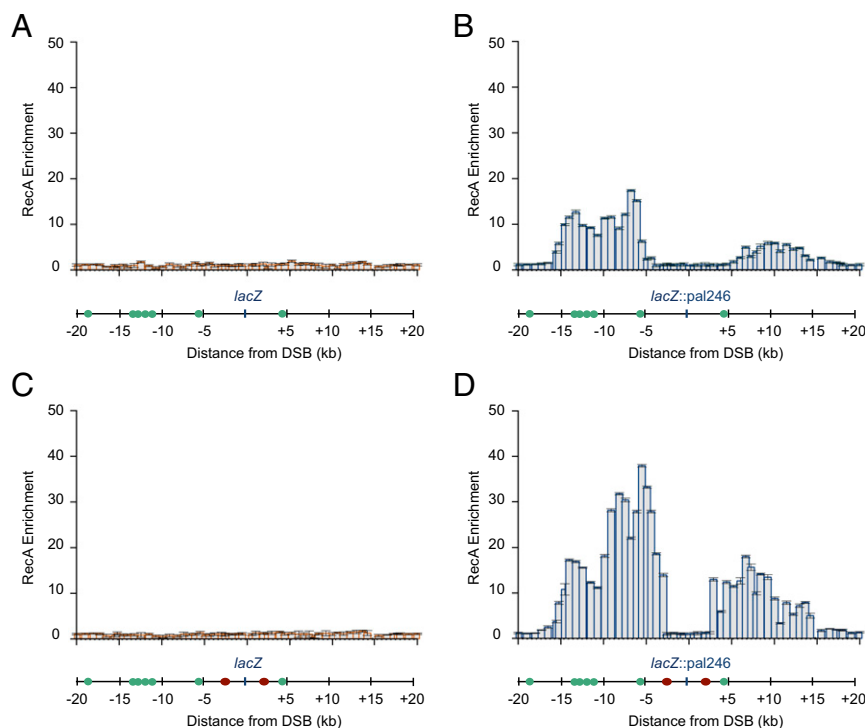
underlying mechanisms giving rise to the binding (18). We reasoned that given the detailed mechanistic understanding of RecBCD in vitro, we could gain a deeper insight into its in vivo functions by developing a mathematical model of RecBCD action that would enable us to estimate the mechanistic parameters of the complex in live cells. Our ChIP data indicate that RecA is indeed loaded on to DNA in a Chi-dependent manner and we have used our mathematical model to infer the parameters of RecBCD action in vivo on a genomic scale. Furthermore, our analysis reveals that DSBR at *lacZ* induces DSBR in the terminus region of the chromosome, an unanticipated observation illuminated by the genomic scale of our data.

## Results

**DSB-Dependent RecA Loading to DNA.** We initially investigated the in vivo binding of RecA at the site of a DSB by ChIP and assayed RecA–DNA interactions by quantitative PCR (qPCR). The DSB was generated by SbcCD-mediated cleavage of a 246-bp interrupted DNA palindrome inserted in the *lacZ* gene (*lacZ*::pal246) (1). In the absence of a DSB, there was no RecA enrichment detected in the 40-kb region surrounding *lacZ* (Fig. 2A). However, following the induction of a DSB there was significant RecA binding detected on both sides of *lacZ* (Fig. 2B). This binding corresponded to the first correctly orientated Chi site on either side of the DSB and spread out over several kilobases of DNA, consistent with the formation of a RecA filament on a single-strand of DNA generated by RecBCD, followed by strand invasion to form a joint molecule. These data suggested that, as expected and consistent with in vitro data, RecA is loaded at the DSB in a Chi-dependent manner (12, 19).

After recognition of the Chi sequence, qPCR analysis detected the binding of RecA to a 30-kb region of DNA surrounding the DSB. Large peaks of RecA enrichment were detected immediately after the first Chi sites on both sides of the DSB, with RecA enrichment decreasing at Chi sites further away from *lacZ* (Fig. 2B). On the origin-proximal side of the break we detected binding of RecA following, not only the first Chi site encountered, but also to subsequent Chi sites. Eighteen-fold RecA enrichment was observed following the Chi site positioned closest to the DSB at a locus on the origin-proximal side and a subsequent peak of 12-fold RecA enrichment was detected at loci positioned ~13 kb origin-proximal to the DSB. This second peak is consistent with the presence of four Chi sites in this region. The origin-distal side of the DSB does not have a second peak, with RecA enrichment plateauing at fivefold enrichment at the sites tested between 7 kb and 13 kb, as expected given the presence of only a single Chi site in this region. The distribution of RecA binding also suggested that the two sides of the DSB might be processed differently, with a higher RecA-enrichment observed at the first Chi position on the origin-proximal side of the break compared with the origin-distal side (Fig. 2B).

**RecA Loading at Synthetic Arrays of Three Chi Sites.** To confirm that RecA was indeed loaded in relation to the recognition of Chi sites by the RecBCD enzyme, we investigated RecA binding in the presence of synthetic arrays of three Chi sites inserted at 3 kb on either side of the DSB. In vitro studies have shown that a single Chi site is recognized by RecBCD with an efficiency of 20–40%, which suggests that following a DSB, a substantial number of RecBCD molecules fail to recognize Chi (20). An efficiency of Chi recognition in vivo similar to that obtained in vitro would explain the observed Chi distribution at *lacZ*::pal246. Previously, arrays of three synthetic Chi sites have been shown to be recognized by RecBCD with an efficiency of 60–80% (21). We reasoned that placing these arrays either side of the DSB would focus a similar proportion of RecA loading closer to *lacZ*. Furthermore, we placed the Chi sites at equal distances (3 kb) on the two sides of the break to increase the symmetry of the reaction. ChIP-qPCR



**Fig. 2.** RecA binding following a DSB at *lacZ::pal246*. ChIP in combination with qPCR revealed that RecA binding to DNA is DSB-dependent and stimulated in response to both endogenous Chi sites (A and B) and artificial Chi arrays (C and D) surrounding the DSB. ChIP-qPCR primer pairs were designed to generate amplicons that span a 40-kb region surrounding the DSB. Chi sites are represented by green circles and triple-Chi arrays by red circles on the map. All Chi sites shown are in the active orientation. The palindrome (*pal246*) is located at position 0 kb on the map (A) Strain DL1777 (*lacZ<sup>+</sup> sbcDC<sup>+</sup>*). (B) Strain DL2859 (*lacZ::pal246 sbcDC<sup>+</sup>*). (C) Strain DL4899 (DL1777 + *lacZY::XXX mhpA::XXX*) (D) Strain DL4900 (DL2859 + *lacZY::XXX mhpA::XXX*). RecA binding at the DSB was normalized to a control site located in the *hycG* gene on the opposite side of the chromosome. Error bars represent the SEM, where  $n = 3$ .

revealed that, in vivo, the triple Chi arrays do indeed stimulate RecA binding closer to the DSB and that binding was enhanced relative to that observed at single endogenous Chi sites (Fig. 2 B and D). Interestingly, the asymmetry in RecA binding to the DNA following a DSB remained, with more RecA bound to the origin-proximal side of the DSB compared with the origin-distal side (Fig. 2D). Furthermore, despite the 38-fold RecA enrichment detected at a locus ~3 kb origin-proximal to the DSB, there was still as much as 15-fold RecA enrichment observed at loci following endogenous Chi sites subsequent to the triple-Chi arrays. This finding confirmed that, like single Chi sites, the triple Chi arrays failed to be recognized in a detectable proportion of the population and that successive Chi sites are required for efficient DSBR.

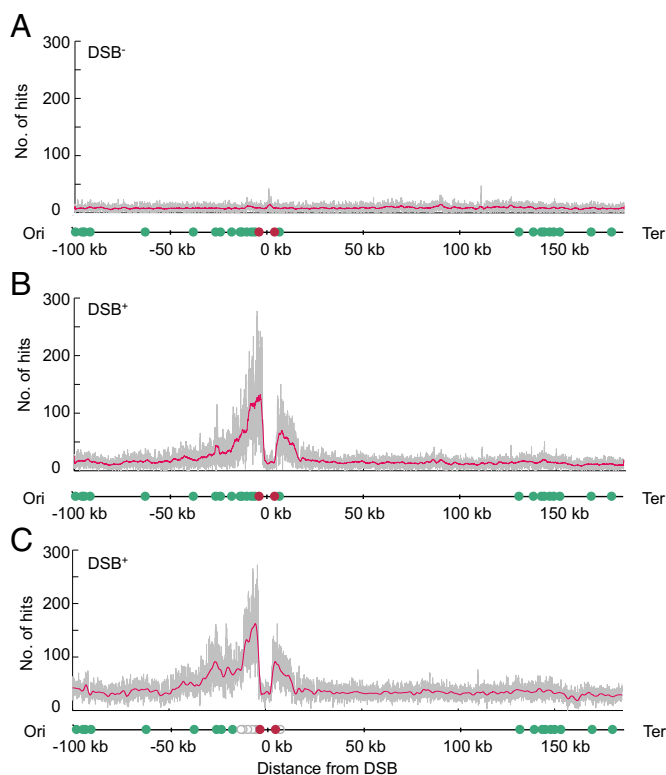
**High-Resolution Analysis of RecA Loading by ChIP-Seq.** To quantify RecA binding to DNA in relation to Chi, ChIP was combined with high-throughput sequencing (ChIP-Seq) to provide a genome-wide analysis of RecA–DNA interactions following a DSB (Fig. 3). These experiments were carried out with the arrays of three Chi sites at 3 kb on either side of the DSB site to focus the reaction at equidistant sites on both sides of the break. RecA was loaded, at the site of a DSB in the *lacZ* gene, in a Chi-dependent manner with approximately twofold more RecA on the origin-proximal side of the DSB compared the origin-distal side (Fig. 3B). This finding is consistent with the results obtained by ChIP-qPCR (Fig. 2) and suggests that the two DNA ends are not equally competent for Chi recognition. Because previous work has shown that SbcCD generates a two-ended break in a RecB mutant (1), we suggest that approximately half of the DSBs arising at the interrupted palindrome are converted from two-ended to one-ended structures by RecBCD action. This could happen if RecBCD traveling on the origin-distal end catches up

with the replication fork and dissociates before recognizing a Chi sequence (Fig. S1).

To further determine the role of subsequent Chi sites during DSBR, we deleted all of the endogenous Chi sites within a 15-kb region on either side of the break, leaving only the triple-Chi arrays positioned 3 kb either side of the *lacZ*. ChIP-Seq analysis revealed that, although the level of enrichment stimulated by the triple-Chi array remained unchanged, RecA binding was decreased in the 15-kb region where the Chi sites had been deleted (Fig. 3C). Strikingly, this decrease was correlated with an increase in RecA binding caused by Chi sites more than 20 kb away from the DSB (Fig. S2). This finding confirmed that RecBCD enzyme complexes that did not act at the array of three Chi sites progress many kilobases further on the DNA until they do recognize a Chi site.

**Inferring the Parameters of RecBCD Activity in Vivo from High-Resolution ChIP-Seq Data Using Mathematical Modeling.** We reasoned that the high spatial resolution afforded by ChIP-Seq data could be exploited to reveal quantitative aspects of the molecular behavior of RecBCD-mediated loading of RecA in living cells. To interpret the quantitative information contained in the high-resolution ChIP-Seq data in terms of the parameters of RecBCD action in vivo, we developed a mathematical model of enzyme action. Our mathematical model is based on the known in vitro properties of RecBCD and its crystal structure (7, 8, 22) and is described in detail in the *SI Appendix* (see also Fig. 1A). The model's main assumptions are the following: Before Chi recognition, the RecBCD complex translocates along the DNA with both the RecD and RecB motors engaged. RecD is the lead motor and the difference in speed of the two motors leads to the accumulation of a ssDNA loop (Loop 1) ahead of RecB that





**Fig. 3.** High-resolution analysis of DSB-dependent RecA loading by ChIP-Seq. (A) ChIP-Seq analysis of the *lacZ* region in the absence of a DSB carried out in strain (DL4899) lacking the 246-bp palindrome. In A–C red circles represent triple-Chi arrays and green circles single Chi sites, all Chi sites shown are in the active orientation. The raw distributions of hits are depicted in gray. To help visualization, the data were filtered using a loess filter (bandwidth 5,700 nucleotides, span 0.02, red curve). The data shown are from a representative example of two biological replicates. (B) ChIP-Seq analysis of the *lacZ* region undergoing DSBR in a strain (DL4900) carrying the 246-bp palindrome located at 0 kb on the map. (C) ChIP-Seq analysis of the *lacZ* region of strain DL5215 demonstrates that removing the endogenous Chi sites within a 30-kb region surrounding the DSB stimulates RecA loading at more distant Chi sites during DSBR. White circles represent endogenous Chi sites that have been deleted.

depends on the motor speed ratio ( $V_B/V_D$ ) and the distance traveled by the enzyme. Chi recognition is stochastic and we denote  $p_{\text{Chi}}$  the probability of Chi recognition. Upon Chi recognition, the RecD motor is disengaged and Loop 1 is converted to Loop/Tail2, which is extended using the RecB motor and RecA is loaded with equal probability across the ssDNA. We denote  $p_{\text{stop}}$  the probability of RecBCD dissociating from DNA or stopping RecA loading. Using these assumptions, we calculated the probability of RecA loading at a genomic position in the vicinity of the DSB depending on the position of the DSB and the Chi sites,  $V_B/V_D$ ,  $p_{\text{Chi}}$ , and  $p_{\text{stop}}$ .

Using a strain in which we had deleted all of the endogenous Chi sites within a 15-kb region either side of the break, leaving only the artificial Chi arrays positioned 3 kb either side of the *lacZ*, we varied the number of Chi sites in the origin-proximal array from one to six. We then compared the mathematical model prediction to the RecA ChIP-Seq data obtained for strains. We observed that as the number of Chi sites in the origin-proximal array was increased, RecA binding close to the DSB was increased relative to the proportion of events involving recognition of Chi sites further away from the break (Fig. 4; a direct comparison between one Chi site and the six Chi array is provided in Fig. S3). RecA loading at Chi sites 40 kb away

from the break was most clearly noticeable in the strain with a single Chi site 3-kb origin-proximal to the break (Fig. 4A). Strikingly, the mathematical model accurately captured the shape of the RecA distribution in all different configurations of Chi positioning with respect to the DSB, indicating that it reflects the main features of RecBCD-mediated end resection (Fig. 4A–F).

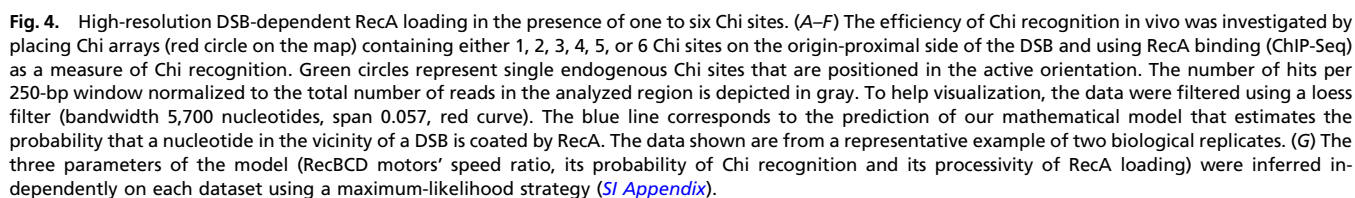
We have used maximum-likelihood estimation to infer the parameters of the mathematical model from the in vivo data (Fig. 4G). Whereas  $p_{\text{Chi}}$  (0.20–0.43) and RecBCD processivity (10 kb) estimates were close to those obtained in vitro, the motor speed ratio,  $V_B/V_D$  (0.94–0.96), was significantly higher than previously reported in vitro (9, 19, 20, 23, 24). Interestingly, we observed that the mathematical model's estimate for  $p_{\text{Chi}}$  decreased as the number of Chi sites in the array was increased (Fig. 4G). As  $p_{\text{Chi}}$  is the probability of recognizing one Chi site, this suggests that when Chi sites are positioned very close together (Chi sites are positioned 10 bp apart in the artificial Chi arrays) they are not recognized independently by RecC. This would lead to an underestimation of  $p_{\text{Chi}}$  in the strains that have multiple Chi sites in the array. Therefore, we focused our interpretation of the data on the strain with only one Chi site positioned 3-kb origin-proximal of the DSB (Fig. 4A). Recent in vitro single-molecule experiments have suggested that there may be two populations of RecBCD molecules each with a different velocity (25). We extended the mathematical model allowing for two populations of RecBCD with distinct  $V_B/V_D$  and  $p_{\text{Chi}}$ . Strikingly, the extended model showed a better fit to the data (SI Appendix, Fig. 7) and had a lower Bayesian Information Criterion score, indicating that this better fit is statistically significant. Maximum-likelihood estimates of the parameters of this extended model indicated two clearly separated populations with 46% of RecBCD with low  $p_{\text{Chi}}$  (0.26) and high  $V_B/V_D$  (0.86) and 54% with higher  $p_{\text{Chi}}$  (0.86) and lower  $V_B/V_D$  (0.58) (SI Appendix).

#### ChIP-Seq Reveals RecA Binding to Other Regions in the Genome, Including DSB-Dependent Binding in the Terminus of the Chromosome.

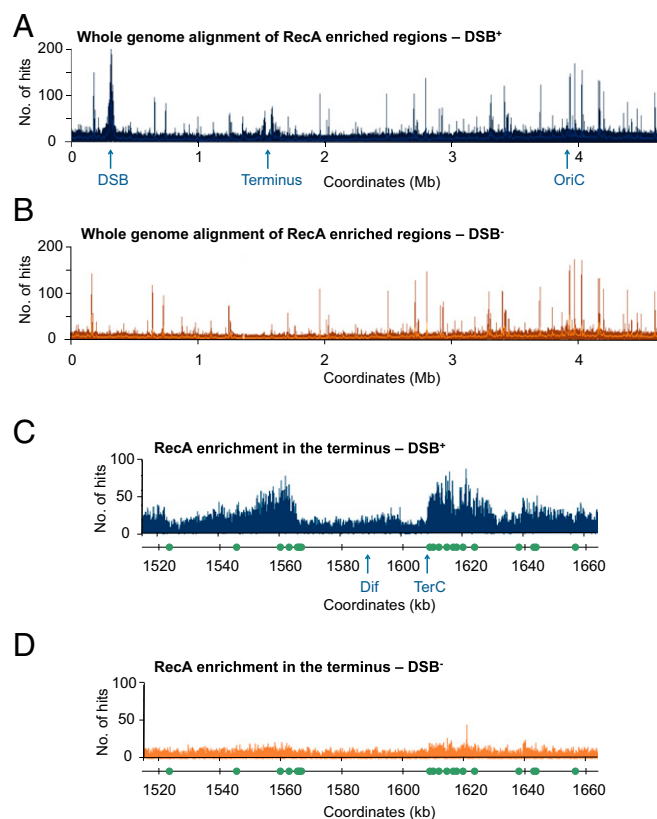
Genome-wide analysis of our dataset revealed DSB-independent RecA binding at distinct loci across the genome (Fig. 5A and B). These loci include the rRNA genes, tRNA genes, and ribosomal protein genes. The positions of these loci of RecA binding were not associated with the positions of Chi sites, suggesting that the RecA binding at these sites is not RecBCD-dependent. ChIP signal at highly transcribed genes has been reported for other proteins, including Smc of *Bacillus subtilis* and SeqA of *E. coli*, and it is unclear whether such a signal is directly related to RecA activity (26, 27). Surprisingly, we observed that the DSB at the *lacZ* locus induced RecA binding in the region of the chromosome involved in the termination of replication (Fig. 5A and C). This RecA binding occurred at positions of Chi sites, characteristic of RecBCD-mediated processing. This finding therefore indicates the presence of additional, indirectly generated, double-strand ends in the region of the chromosome containing *dif*, the site responsible for the resolution of chromosome dimers by XerCD site-specific recombination (28).

#### Discussion

**RecA Protein Binding to a DNA DSB in Vivo Is Determined by the Chi Sites in the Region Surrounding the Break.** We have used a system that accurately introduces a single site-specific DNA DSB into one copy of the replicated *E. coli* chromosome. The system uses the fact that a 246-bp interrupted palindrome is cleaved by the SbcCD enzyme at the site of a DNA hairpin structure formed on only one of the replicated chromosomal copies (1). We predict the hairpin to be formed on the lagging-strand template because of its single-stranded nature and that under our growth conditions, repair occurs efficiently, presumably using the uncleaved sister chromosome as a template (4, 17). Using ChIP in *E. coli* with antibodies against the RecA protein, we investigated the behavior of this protein as it is engaged in repairing this DSB. We



**Inference of the Parameters of RecBCD Action in Vivo.** We have coupled quantitative genomic analysis of RecA binding with mathematical modeling of the RecBCD complex in its loading of



**Fig. 5.** Genomic distribution of RecA binding reveals DSB-dependent RecA binding at the terminus. (A and B) Genome-wide RecA binding. DSB-independent RecA binding was observed at distinct loci across the genome, these loci predominantly correspond the ORFs of genes that encode rRNA, tRNA, and ribosomal proteins. In addition, DSB-dependent binding was observed in the terminus region. (A) Strain DL4900 (*lacZ::pal246 sbcDC<sup>+</sup>*) RecA-ChIP; (B) strain DL4899 (*lacZ::pal246 sbcDC<sup>-</sup>*) RecA-ChIP. (C and D) Zoom-in showing RecA binding in the terminus region of the chromosome. The distribution of RecA binding strongly correlates with the position of correctly oriented Chi sites in the region, indicating a DSB occurs in the terminus region that is repaired by RecBCD-mediated HR. The *dif* site is located at 1,588,774 bp and *terC* is located at 1,607,181 bp. No other *ter* sites are located within this region. (C) Strain DL4900 (*lacZ::pal246 sbcDC<sup>+</sup>*). (D) Strain DL4899 (*lacZ::pal246 sbcDC<sup>-</sup>*). Green circles represent single endogenous Chi sites that are positioned in the active orientation. The raw distributions of hits are depicted in blue and orange.

RecA to infer, based on the assumptions of the mathematical model, the molecular parameters of RecBCD action in live cells. Initial analysis using a mathematical model with one mode of action of RecBCD led to estimates of the probability of Chi recognition and processivity values similar to those that had previously been measured in vitro (19, 20). However, we observed that the inferred ratio of the two motors' speed ( $V_B/V_D = 0.94\text{--}0.96$ ) was significantly higher than the value of 0.6 observed in vitro from studies of mutant enzymes defective for the helicase motors of RecB or RecD (25) and from evaluation of the average rate of Loop 1 formation relative to total unwinding by the wild-type enzyme (9). We calculated (*SI Appendix*, Eq. [2]) that a ratio of 0.95 would result in the production of a single-strand loop before Chi (Loop 1) of 3 kb for a Chi site positioned 60 kb from the break. In contrast, a  $V_B/V_D$  of 0.6 similar to that reported in vitro would result in an extremely long loop of 40 kb in vivo. These differences in  $V_B/V_D$  might be because of differences in RecBCD activity in vitro and in vivo. However, a wide distribution in the values of  $V_B/V_D$  has been observed in vitro (9), and two broad populations of RecBCD molecules with different velocities have

been reported (25). When we extended the mathematical model to explore the possibility of two RecBCD populations, each with a different mode of action, we found that a two-population model was supported by the data. Assuming the existence of two populations could be an oversimplification and we cannot rule out more complex RecBCD populations. However, it is interesting to note that under this two-population model, the inferred parameters show a sharp contrast between molecules with a low probability of Chi recognition associated with a high motor speed ratio and molecules with a higher  $p_{\text{Chi}}$  and lower  $V_B/V_D$ . Both these combinations of parameters will result in approximately the same average length of Loop 1 given the average density of Chi sites on the genome (see calculation in *SI Appendix*). RecBCD complexes with low  $p_{\text{Chi}}$  may have to travel very far before Chi recognition but because of their high  $V_B/V_D$  they accumulate a relatively short Loop 1. In contrast, RecBCD molecules with high  $p_{\text{Chi}}$ , which will recognize Chi motifs close to the break, accumulate a longer Loop 1. This trade-off may indicate that controlling the size of Loop 1 has important consequences for RecBCD function.

The parameter estimates obtained here need to be interpreted within the assumptions of the mathematical model. For example, we have assumed that the whole single-stranded region generated by RecBCD is covered equally well by RecA. However, if only part of it is covered by RecA or if RecA binding extends into the adjacent double-stranded region, the inference of  $p_{\text{stop}}$  and its interpretation would be affected. For example, if RecBCD continues unwinding after ceasing RecA loading, this part of the single strand would not be detected in the experimental assay and RecBCD processivity would be underestimated. Therefore, the  $p_{\text{stop}}$  inferred here is to be understood as an "effective processivity of RecA loading by RecBCD," which is the combination of its DNA unwinding and RecA loading activities. Similarly, the estimation of  $V_B/V_D$  could be affected if RecA is not loaded with the same probability across the Chi-proximal region of the ssDNA.

#### A DSB at *lacZ* Induces DSBR in the Terminus Region of the Chromosome.

In our system, we induce a DSB at the site of an interrupted DNA palindrome inserted at the *lacZ* locus, which lies about half way between the single origin of replication and the terminus. Our genomic analysis has revealed that DSBR (characterized by Chi-correlated RecA binding) in the terminus region surrounding *dif*, at a distance of over 1 Mb from *lacZ*. This RecA binding indicates that following a DSB at *lacZ*, dsDNA ends are generated in the region containing the *dif* site, which is required for the resolution of chromosome dimers by XerCD (28). The signal in this region is significantly lower than at *lacZ*, suggesting that these double-stranded ends only appear in a subpopulation of cells. However, it is currently unclear how a break at *lacZ* causes breakage in the terminus region. The RecA bound is approximately symmetrically distributed on the two sides of the *dif*-containing region. The observation that strains undergoing these breaks are fully viable (17) leads us to believe that unbroken sister chromosomal DNA in the *dif*-containing region must also be present to facilitate efficient repair. Whether this RecA binding implies the presence of two-ended breaks or of two, equally frequent, single-ended breaks remains to be determined.

Interestingly, the existence of double-strand ends in the terminus region of the chromosome has been hypothesized previously. Kogoma has proposed that recombination-dependent DNA replication may be responsible for induced stable DNA replication, which can be initiated at a sequence known as *oriM2* in the terminus region (29). However, *oriM2* has not been mapped accurately and it is not known whether the DSB that we observe relates to this origin. Several other results, such as the existence of terminal recombination (30–32) and the striking replication profile of a *recB* mutant (33), indicate that the terminus region of the chromosome presents an area of importance



to recombination. However, there are very likely several different reactions taking place. Our observation of a DSB event close to *dif* induced by DSB at *lacZ* provides a clear physical demonstration of one such interaction in this region.

## Experimental Procedures

**Bacterial Strains and Growth.** All strains are derivatives of *E. coli* K12 MG1655 (34) and are listed in Table S1. Cells were grown in M9 minimal media supplemented with 0.2% casamino acids, 0.5% glucose, 5  $\mu$ M  $\text{CaCl}_2$ , and 1 mM  $\text{MgSO}_4$  at 37 °C. Mutations were introduced by P1 transduction or plasmid-mediated gene replacement (PMGR) (35–38) using the plasmids described in Table S2. All primers used for cloning and genotyping are detailed in Table S3.

**ChIP Sample Preparation.** All ChIP experiments were performed with  $\sim 5 \times 10^8$  cells growing in exponential growth phase ( $\text{OD}_{600\text{nm}}$  0.2–0.25). RecA–DNA interactions were chemically cross-linked with formaldehyde (Sigma-Aldrich; final concentration 1%) for 10 min at 22.5 °C. Cross-linking was quenched by the addition of glycine (Sigma-Aldrich; final concentration 0.5 M). Cells were collected by centrifugation and washed three times in ice-cold 1 $\times$  PBS. The pellet was then resuspended in 250  $\mu$ L ChIP buffer [200 mM Tris-HCl (pH 8.0), 600 mM NaCl 4% (vol/vol) Triton X, Complete protease inhibitor mixture EDTA-free (Roche)]. Sonication of cross-linked samples was performed using the Diagenode Bioruptor at 30-s intervals for 10 min at high amplitude. After sonication, 350  $\mu$ L of ChIP buffer was added to each sample, the samples were mixed by gentle pipetting and 100  $\mu$ L of each lysate was removed and stored as “input.” Immunoprecipitation was performed overnight at 4 °C using 1/100 anti-RecA antibody (Abcam, ab63797). IP samples were then incubated with Protein G Dynabeads (Life Technologies) for 2 h at room temperature. All samples were washed three times with 1 $\times$  PBS + 0.02% Tween-20 before resuspending the Protein G dynabeads in 200  $\mu$ L of TE buffer [10 mM Tris (pH 7.4), 1 mM EDTA] + 1% SDS. Next, 100  $\mu$ L of TE buffer was added to the input samples and all samples were then incubated at 65 °C for 10 h to reverse formaldehyde cross-links. DNA was isolated using the MinElute PCR purification kit (Qiagen). DNA was eluted in 50  $\mu$ L of TE buffer using a two-step elution. Samples were stored at –20 °C.

**Library Preparation for High-Throughput Sequencing.** Input and ChIP samples were processed following New England Biolabs’ protocol from the NEBNext ChIP-Seq library preparation kit. Briefly, 200 ng of input and ChIP-enriched DNA was subject to end repair to fill in ssDNA overhangs, remove 3’ phosphates and 5’ phosphorylate the sheared DNA. Klenow exo- was used to adenylate the 3’ ends of the DNA and NEXTflex DNA barcodes (Bioo Scientific) were ligated using T4 DNA ligase. After each step, the DNA was purified using the Qiagen MinElute PCR purification kit according to the manufacturer’s instructions. After adaptor ligation, the adaptor-modified DNA fragments were enriched by PCR using primers corresponding to the beginning of each adaptor. Finally, agarose gel electrophoresis was used to size select adaptor-ligated DNA with an average size of  $\sim 275$  bp. All samples were quantified on a Bioanalyzer (Agilent) before being sequenced on the Illumina HiSeq. 2000. RecA binds to both ssDNA and dsDNA in presynaptic and postsynaptic complexes (39–41). It was previously believed that ssDNA could not be detected by ChIP-Seq. However, several studies have recently shown that this is not the case; ssDNA is rendered double-stranded during the library preparation process through the formation of DNA hairpins that arise as a result of regions of microhomology (42–44). This allows the DNA to be amplified and detected by ChIP-Seq. These findings are consistent with our data, which shows a similar pattern of RecA binding detected using both qPCR and high-throughput sequencing.

**ChIP-Seq Data Analysis.** For ChIP-Seq analysis, 50-bp single-end reads were mapped to the *E. coli* K12 MG1655 (NC000913.3) (34) genome using Novoalign v2.07 ([www.novocraft.com](http://www.novocraft.com)). Novoalign uses the Needleman–Wunsch algorithm to determine the optimal alignment of reads. Before mapping, the 3’ adaptor sequences were removed using fastx\_clipper and the data collapsed using fastx\_collapser to remove identical sequence reads ([hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)). The preparation of ChIP-Seq libraries requires a PCR of the adaptor ligated DNA. This can result in PCR duplication of certain DNA fragments. Removing duplicates mitigates the effects of PCR amplification bias so that regions of the genome don’t appear more enriched than they actually are. The ChIP-Seq datasets in this study contained  $\sim 4\%$  PCR duplicates and these were discarded. The data were also plotted without removing these duplicates and revealed that the trend in RecA binding was unchanged. Sequences were mapped with default parameters, allowing for a maximum of one mismatch per read (novoalign -f DL4900\_IP.fasta -d DL4900\_genome.nix -r Random > DL4900\_IP.sam). To report reads that have multiple alignment loci, we specified the –r parameter as either –r Random or –r None. In the first case Novoalign chooses a single alignment location at random among all of the alignment results; in the second case, only the reads that map to a single genomic location are aligned ([www.novocraft.com](http://www.novocraft.com)). PyReadCounters was used to calculate the overlap between aligned reads and *E. coli* genomic features (45). The distribution of reads along the *E. coli* genome was visualized using the Integrated Genome Browser (46). Full details of all scripts are available upon request.

**Identification of RecA-Binding at the DSB.** Because of the specific mechanism of RecBCD-mediated RecA loading observed around the DSB, classic peak-calling algorithms such as MACS (47) failed to recapitulate the RecA binding at this site. This is because RecA loading at a DSB is the result of a complex dynamic process that cannot be described as a simple binding event. This suggests that, as has been observed for other datasets (48), the shape of the peaks may carry important information. In particular, we reasoned that given the high spatial resolution of ChIP-Seq data, the position and shape of the peaks observed at Chi sites could give us quantitative information about the mechanism of RecBCD-mediated DSB repair in vivo. Therefore, we developed a mathematical model of RecBCD-dependent RecA loading to evaluate the probability that a nucleotide in the vicinity of a DSB is coated by the RecA protein. We then used maximum-likelihood estimation to extract the parameters of this model from the dataset. This mathematical model and the associated data analysis are described in detail in the SI Appendix.

**qPCR.** All real-time qPCR reactions were carried out in 15- $\mu$ L volumes in the MX3000P qPCR machine (Agilent) using the Brilliant II SYBR Green qPCR master mix (Agilent). The temperature profile for all assays was 95 °C for 10 min followed by 40 cycles of 95 °C for 20 s and 60 °C for 60 s. All reactions were repeated in triplicate and the formation of PCR products of the correct lengths was confirmed by agarose gel electrophoresis. A full list of primers used for qPCR is given in Table S4. Assay performance was checked by standard curve for all assays. Data were exported from the MxPro software to Microsoft Excel for analysis. The melting temperature of the qPCR primers was calculated by the manufacturer (MWG Biotech).

**ACKNOWLEDGMENTS.** We thank Dr. N. Molina and Dr. G. Sanguinetti for advice on data analysis; Dr. Sander Granneman for advice on ChIP-Seq analysis; Dr. Ralph Hector for advice on ChIP-Seq library preparation; and Dr. M. White for the construction of pDL4690. This research has been supported by an Medical Research Council studentship and a Medical Research Council Centenary Award (to C.A.C.); a Darwin Trust of Edinburgh postgraduate studentship (to M.F.); Marie Curie Fellowship POF-GA-2009-254082 – DRIBAC (to M.E.K.); European Research Council Advanced Grant RULE-320823 (to V.D.); and a Medical Research Council programme Grant G0901622 (to D.R.F.L.).

1. Eykelenboom JK, Blackwood JK, Okely E, Leach DRF (2008) SbcCD causes a double-strand break at a DNA palindrome in the *Escherichia coli* chromosome. *Mol Cell* 29(5):644–651.
2. Cromie GA, Connolly JC, Leach DR (2001) Recombination at double-strand breaks and DNA ends: Conserved mechanisms from phage to humans. *Mol Cell* 8(6):1163–1174.
3. Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehauer WM (1994) Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev* 58(3):401–465.
4. Mawer JS, Leach DR (2014) Branch migration prevents DNA loss during double-strand break repair. *PLoS Genet* 10(8):e1004485.
5. Connolly B, et al. (1991) Resolution of Holliday junctions in vitro requires the *Escherichia coli* *ruvC* gene product. *Proc Natl Acad Sci USA* 88(14):6063–6067.
6. Connolly B, West SC (1990) Genetic recombination in *Escherichia coli*: Holliday junctions made by RecA protein are resolved by fractionated cell-free extracts. *Proc Natl Acad Sci USA* 87(21):8476–8480.
7. Dillingham MS, Kowalczykowski SC (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* 72(4):642–671.
8. Smith GR (2012) How RecBCD enzyme and Chi promote DNA break repair and recombination: A molecular biologist’s view. *Microbiol Mol Biol Rev* 76(2):217–228.
9. Taylor AF, Smith GR (2003) RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature* 423(6942):889–893.
10. Handa N, et al. (2012) Molecular determinants responsible for recognition of the single-stranded DNA regulatory sequence,  $\chi$ , by RecBCD enzyme. *Proc Natl Acad Sci USA* 109(23):8901–8906.
11. Dixon DA, Kowalczykowski SC (1995) Role of the *Escherichia coli* recombination hotspot, Chi, in RecABCD-dependent homologous pairing. *J Biol Chem* 270(27):16360–16370.
12. Anderson DG, Kowalczykowski SC (1998) SSB protein controls RecBCD enzyme nuclease activity during unwinding: A new role for looped intermediates. *J Mol Biol* 282(2):275–285.
13. Ponticelli AS, Schultz DW, Taylor AF, Smith GR (1985) Chi-dependent DNA strand cleavage by RecBC enzyme. *Cell* 41(1):145–151.



14. Taylor AF, Schultz DW, Ponticelli AS, Smith GR (1985) RecBC enzyme nicking at Chi sites during DNA unwinding: Location and orientation-dependence of the cutting. *Cell* 41(1):153–163.
15. Anderson DG, Kowalczykowski SC (1997) The recombination hot spot Chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes Dev* 11(5):571–581.
16. Taylor AF, Smith GR (1999) Regulation of homologous recombination: Chi inactivates RecBCD enzyme by disassembly of the three subunits. *Genes Dev* 13(7):890–900.
17. Darmon E, Eykelenboom JK, Lopez-Vernaza MA, White MA, Leach DR (2014) Repair on the go: *E. coli* maintains a high proliferation rate while repairing a chronic DNA double-strand break. *PLoS One* 9(10):e110784.
18. Bailey T, et al. (2013) Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLOS Comput Biol* 9(11):e1003326.
19. Dixon DA, Kowalczykowski SC (1993) The recombination hotspot Chi is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. *Cell* 73(1):87–96.
20. Taylor AF, Smith GR (1992) RecBCD enzyme is altered upon cutting DNA at a chi recombination hotspot. *Proc Natl Acad Sci USA* 89(12):5226–5230.
21. Spies M, et al. (2003) A molecular throttle: The recombination hotspot Chi controls DNA translocation by the RecBCD helicase. *Cell* 114(5):647–654.
22. Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* 432(7014):187–193.
23. Bianco PR, et al. (2001) Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* 409(6818):374–378.
24. Taylor A, Smith GR (1980) Unwinding and rewinding of DNA by the RecBC enzyme. *Cell* 22(2 Pt 2):447–457.
25. Liu B, Baskin RJ, Kowalczykowski SC (2013) DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate. *Nature* 500(7463):482–485.
26. Waldminghaus T, Skarstad K (2010) ChIP on Chip: Surprising results are often artifacts. *BMC Genomics* 11:414.
27. Gruber S, Errington J (2009) Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in *B. subtilis*. *Cell* 137(4):685–696.
28. Barre FX, et al. (2001) Circles: The replication-recombination-chromosome segregation connection. *Proc Natl Acad Sci USA* 98(15):8189–8195.
29. Kogoma T (1997) Stable DNA replication: Interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* 61(2):212–238.
30. Corre J, Cornet F, Patte J, Louarn JM (1997) Unraveling a region-specific hyper-recombination phenomenon: Genetic control and modalities of terminal recombination in *Escherichia coli*. *Genetics* 147(3):979–989.
31. Horiuchi T, Fujimura Y, Nishitani H, Kobayashi T, Hidaka M (1994) The DNA replication fork blocked at the Ter site may be an entrance for the RecBCD enzyme into duplex DNA. *J Bacteriol* 176(15):4656–4663.
32. Wendel BM, Courcelle CT, Courcelle J (2014) Completion of DNA replication in *Escherichia coli*. *Proc Natl Acad Sci USA* 111(46):16454–16459.
33. Rudolph CJ, Upton AL, Stockum A, Nieduszynski CA, Lloyd RG (2013) Avoiding chromosome pathology when replication forks collide. *Nature* 500(7464):608–611.
34. Blattner FR, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277(5331):1453–1462.
35. Link AJ, Phillips D, Church GM (1997) Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: Application to open reading frame characterization. *J Bacteriol* 179(20):6228–6237.
36. Merlin C, McAteer S, Masters M (2002) Tools for characterization of *Escherichia coli* genes of unknown function. *J Bacteriol* 184(16):4573–4581.
37. Darmon E, et al. (2007) SbcCD regulation and localization in *Escherichia coli*. *J Bacteriol* 189(18):6686–6694.
38. White MA, Eykelenboom JK, Lopez-Vernaza MA, Wilson E, Leach DR (2008) Non-random segregation of sister chromosomes in *Escherichia coli*. *Nature* 455(7217):1248–1250.
39. Chen Z, Yang H, Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453(7194):489–494.
40. Galletto R, Amitani I, Baskin RJ, Kowalczykowski SC (2006) Direct observation of individual RecA filaments assembling on single DNA molecules. *Nature* 443(7113):875–878.
41. Pugh BF, Cox MM (1987) Stable binding of recA protein to duplex DNA. Unraveling a paradox. *J Biol Chem* 262(3):1326–1336.
42. Croucher NJ, et al. (2009) A simple method for directional transcriptome sequencing using Illumina technology. *Nucleic Acids Res* 37(22):e148.
43. Khil PP, Smagulova F, Brick KM, Camerini-Otero RD, Petukhova GV (2012) Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. *Genome Res* 22(5):957–965.
44. Yamane A, et al. (2013) RPA accumulation during class switch recombination represents 5′-3′ DNA-end resection during the S-G2/M phase of the cell cycle. *Cell Reports* 3(1):138–147.
45. Webb S, Hector RD, Kudla G, Granneman S (2014) PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol* 15(1):R8.
46. Nicol JW, Helt GA, Blanchard SG, Jr, Raja A, Loraine AE (2009) The Integrated Genome Browser: Free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25(20):2730–2731.
47. Zhang Y, et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
48. Schweikert G, Cseke B, Clouaire T, Bird A, Sanguinetti G (2013) MMDiff: Quantitative testing for shape changes in ChIP-Seq data sets. *BMC Genomics* 14:826.